

Westphal, Andrea; Gronostaj, Anna; Vock, Miriam; Emmrich, Rico; Harych, Peter  
**Differenzierung im gymnasialen Mathematik und Deutschunterricht - vor allem bei guten Diagnostiker/innen und in heterogenen Klassen?**

*Zeitschrift für Pädagogik 62 (2016) 1, S. 131-148*



**Quellenangabe/ Reference:**

Westphal, Andrea; Gronostaj, Anna; Vock, Miriam; Emmrich, Rico; Harych, Peter: Differenzierung im gymnasialen Mathematik und Deutschunterricht - vor allem bei guten Diagnostiker/innen und in heterogenen Klassen? - In: Zeitschrift für Pädagogik 62 (2016) 1, S. 131-148 - URN: urn:nbn:de:0111-pedocs-167092 - DOI: 10.25656/01:16709

<https://nbn-resolving.org/urn:nbn:de:0111-pedocs-167092>

<https://doi.org/10.25656/01:16709>

in Kooperation mit / in cooperation with:

**BELTZ JUVENTA**

<http://www.juventa.de>

**Nutzungsbedingungen**

Gewährt wird ein nicht exklusives, nicht übertragbares, persönliches und beschränktes Recht auf Nutzung dieses Dokuments. Dieses Dokument ist ausschließlich für den persönlichen, nicht-kommerziellen Gebrauch bestimmt. Die Nutzung stellt keine Übertragung des Eigentumsrechts an diesem Dokument dar und gilt vorbehaltlich der folgenden Einschränkungen: Auf sämtlichen Kopien dieses Dokuments müssen alle Urheberrechtshinweise und sonstigen Hinweise auf gesetzlichen Schutz beibehalten werden. Sie dürfen dieses Dokument nicht in irgendeiner Weise abändern, noch dürfen Sie dieses Dokument für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, aufführen, vertreiben oder anderweitig nutzen. Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

**Terms of use**

We grant a non-exclusive, non-transferable, individual and limited right to using this document.

This document is solely intended for your personal, non-commercial use. Use of this document does not include any transfer of property rights and it is conditional to the following limitations: All of the copies of this documents must retain all copyright information and other information regarding legal protection. You are not allowed to alter this document in any way, to copy it for public or commercial purposes, to exhibit the document in public, to perform, distribute or otherwise use the document in public.

By using this particular document, you accept the above-stated conditions of use.

**Kontakt / Contact:**

peDOCS  
DIPF | Leibniz-Institut für Bildungsforschung und Bildungsinformation  
Informationszentrum (IZ) Bildung  
E-Mail: [pedocs@dipt.de](mailto:pedocs@dipt.de)  
Internet: [www.pedocs.de](http://www.pedocs.de)

# ZEITSCHRIFT FÜR PÄDAGOGIK

Heft 1

Januar/Februar 2016

## ■ *Thementeil*

### **Raum und Räumlichkeit in der erziehungswissenschaftlichen Forschung**

## ■ *Allgemeiner Teil*

Fallarbeit als Angebot – fallbasiertes Lernen als Nutzung: Empirische Ergebnisse zur kognitiven Belastung, Motivation und Emotionen bei der Arbeit mit Unterrichtsfällen

„Und sie bewegt sich doch!“ – Wie spezifische Lerngelegenheiten die bildungswissenschaftlichen Kompetenzen von Lehramtsstudierenden fördern können

Differenzierung im gymnasialen Mathematik- und Deutschunterricht – vor allem bei guten Diagnostiker/innen und in heterogenen Klassen?

## Inhaltsverzeichnis

### *Thementeil: Raum und Räumlichkeit in der erziehungswissenschaftlichen Forschung*

*Fabian Kessl/Sabine Reh/Edith Glaser/Werner Thole*

Raum und Räumlichkeit in der erziehungswissenschaftlichen Forschung.

Eine Einführung ..... 1

*Fabian Kessl*

Erziehungswissenschaftliche Forschung zu Raum und Räumlichkeit.

Eine Verortung des Thementeils „Raum und Räumlichkeit

in der erziehungswissenschaftlichen Forschung“ ..... 5

*Sebastian Dirks/Caroline Fritsche/Maike Lippelt/Christian Reutlinger*

Zur pädagogischen Herstellung städtischer Räume zwischen Ort

und Klient\*in. Empirische Einblicke und theoretische Rückschlüsse ..... 20

*Sabine Andresen/Susann Fegter/Nora Iranee/Elena Bütow*

Doing Urban Family: Städtischer Raum und elterliche Perspektive

auf Kindheit ..... 34

*Nils Berkemeyer/Björn Hermstein/Veronika Maniti*

Auf dem Weg zu einer raumorientierten Schulsystemforschung.

Was können raumsensible Sozialtheorien für empirische Analysen

regionaler Bildungsdisparitäten leisten? ..... 48

*Jeanette Böhme/Viktoria Flasche/Ina Herrmann*

Die Territorialisierung des (Schul-)Pädagogischen im urbanen Wandel.

Ein Forschungsbeitrag aus der Pädagogischen Morphologie ..... 62

## *Deutscher Bildungsserver*

Linktipps zum Thema „Raum und Räumlichkeit in der erziehungswissenschaftlichen Forschung“ .....	79
--	----

## **Allgemeiner Teil**

*Marcus Syring/Thorsten Bohl/Marc Kleinknecht/Sebastian Kuntze/*

*Markus Rehm/Jürgen Schneider*

Fallarbeit als Angebot – fallbasiertes Lernen als Nutzung:

Empirische Ergebnisse zur kognitiven Belastung, Motivation

und Emotionen bei der Arbeit mit Unterrichtsfällen ..... 86

*Wilfried Plöger/Daniel Scholl/Andreas Seifert*

„Und sie bewegt sich doch!“ – Wie spezifische Lerngelegenheiten

die bildungswissenschaftlichen Kompetenzen von Lehramtsstudierenden

fördern können ..... 109

*Andrea Westphal/Anna Gronostaj/Miriam Vock/Rico Emmrich/*

*Peter Harych*

Differenzierung im gymnasialen Mathematik- und Deutschunterricht –

vor allem bei guten Diagnostiker/innen und in heterogenen Klassen? ..... 131

## **Besprechungen**

*Ulrike Greiner*

Ewald Terhart/Hedda Bennewitz/Martin Rothland (Hrsg.):

Handbuch der Forschung zum Lehrerberuf (2. überarbeitete

und erweiterte Auflage) ..... 149

*Beat Manz*

Claude Mouchet/Raymond Bénévent: Von Freinet zu Freud:

Die institutionelle Pädagogik von Fernand Oury ..... 153

## **Dokumentation**

Pädagogische Neuerscheinungen ..... 156

Impressum ..... U3

## Table of Contents

### *Topic: Space and Spatiality in Educational Research*

<i>Fabian Kessl/Sabine Reh/Edith Glaser/Werner Thole</i> Space and Spatiality in Educational Research. An introduction .....	1
<i>Fabian Kessl</i> Educational Research on Space and Spatiality – A localization of the topic “Space and Spatiality in Educational Research” .....	5
<i>Sebastian Dirks/Caroline Fritsche/Maike Lippelt/Christian Reutlinger</i> On the Pedagogical Construction of Urban Spaces between Location and Client – Empirical insights and theoretical conclusions .....	20
<i>Sabine Andresen/Susann Fegter/Nora Iranee/Elena Bütow</i> Doing Urban Family: Urban space and parental perspectives on childhood .....	34
<i>Nils Berkemeyer/Björn Hermstein/Veronika Maniti</i> On the Way to Spatially Oriented Research on the School System – In what way can spatially aware social theories contribute to empirical analyses of regional educational disparities? .....	48
<i>Jeanette Böhme/Viktoria Flasche/Ina Herrmann</i> The Territorialization of the (School-) Pedagogical in the Context of Urban Change – A research contribution by pedagogical morphology .....	62
<i>Deutscher Bildungsserver</i> Tips of links relating to the topic of “Space and Spatiality in Educational Research” .....	79

### *Contributions*

<i>Marcus Syring/Thorsten Bohl/Marc Kleinknecht/Sebastian Kuntze/ Markus Rehm/Jürgen Schneider</i> Casework as Offer – Case-based Learning as Use: Empirical results on cognitive stress, motivation, and emotions when working with teaching cases .....	86
--	----

*Wilfried Plöger/Daniel Scholl/Andreas Seifert*  
“And yet it moves!” – How specific learning opportunities  
can promote educational-scientific competencies of student teachers ..... 109

*Andrea Westphal/Anna Gronostaj/Miriam Vock/Rico Emmrich/  
Peter Harych*  
Differentiation in Mathematics and in German Lessons  
on Grammar School Level – Is it to be found mainly among  
good diagnosticians and in heterogeneous classes? ..... 131

Book Reviews ..... 149

New Books ..... 156

Impressum ..... U3

*Andrea Westphal/Anna Gronostaj/Miriam Vock/Rico Emmrich/Peter Harych*

# Differenzierung im gymnasialen Mathematik- und Deutschunterricht – vor allem bei guten Diagnostiker/innen und in heterogenen Klassen?

**Zusammenfassung:** Das Anliegen der vorliegenden Studie ist es, zu prüfen, inwiefern eine differenzierende Unterrichtspraxis in den Fächern Mathematik und Deutsch im Zusammenhang mit der Wahrnehmung von Leistungsheterogenität und der Diagnosegenauigkeit von Lehrkräften steht. Der Untersuchung lagen Leistungsdaten aus den Vergleichsarbeiten (VERA-8) und Lehrerurteile über  $N = 1803$  Schüler/innen der achten Jahrgangsstufe an Gymnasien im Land Brandenburg zugrunde sowie Schülereinschätzungen zur Differenzierung im Unterricht. Hierarchische Regressionsanalysen zeigen, dass Mathematiklehrkräfte aus Sicht der Schüler/innen umso stärker differenzierend unterrichten, je besser sie die Leistungsstände der Schüler/innen einschätzen können. Im Fach Deutsch findet differenzierender Unterricht in geringerem Umfang statt und es zeigen sich keine Zusammenhänge mit der Diagnosegenauigkeit.

**Schlagworte:** Binnendifferenzierung, Diagnosekompetenzen, Leistungsheterogenität, Unterrichtsqualität, Lehrerprofessionalisierung

## 1. Theoretischer Hintergrund

Groß angelegte Leistungsstudien wie PISA zeigen, dass in Deutschland innerhalb ein und derselben Schulform Schüler/innen mit ganz unterschiedlichen Leistungsständen lernen (Hohn, Schiepe-Tiska, Sälzer & Artelt, 2013; Sälzer, Reiss, Schiepe-Tiska, Prenzel & Heinze, 2013). Auch innerhalb einer Schulklasse können sich die Leistungsstände der Schüler/innen deutlich unterscheiden (Gröhlich, Scharenberg & Bos, 2009). Damit alle Schüler/innen trotz dieser heterogenen Leistungsvoraussetzungen optimale Ergebnisse erzielen oder gewisse Mindeststandards erreichen, ist eine Anpassung des Unterrichts an die Lernvoraussetzungen der Schüler/innen erforderlich. Eine wesentliche Möglichkeit hierfür ist, unterschiedliche Lerngelegenheiten bereitzustellen, was in Deutschland allerdings nur selten realisiert wird (z. B. Lankes & Carstensen, 2007; Nieder & Frühauf, 2012; Schrader & Helmke, 2008; Schiepe-Tiska et al., 2013). Um eine solche differenzierende Unterrichtspraxis stärker zu etablieren, bedarf es einer Analyse der Bedingungen, unter denen Lehrkräfte differenzierend unterrichten. In der vorliegenden Studie prüfen wir deshalb, ob Lehrkräfte leistungsdifferenzierende Aufgaben – als eine mögliche Form der Binnendifferenzierung – in stärkerem Umfang einsetzen, wenn sie ihre Klasse als sehr leistungsheterogen wahrnehmen. Außerdem untersu-

chen wir, ob Lehrkräfte in stärkerem Maße binnendifferenzierend unterrichten, wenn sie über höhere diagnostische Kompetenzen verfügen.

### 1.1 Binnendifferenzierung statt Orientierung am Klassendurchschnitt

Die Adaptivität des Unterrichts wird als ein zentrales Merkmal der Unterrichtsqualität angesehen (Ditton, 2000, 2002; Slavin, 1994). Grundlegend ist die Idee, dass der größte Kompetenzzuwachs dann erreicht werden kann, wenn Lern- und Unterstützungsangebote so auf die unterschiedlichen Voraussetzungen der Schüler/innen abgestimmt sind, dass diese sie leicht aufnehmen und verarbeiten können und weder unterfordert noch überfordert sind. Eine Passung des Unterrichts kann insbesondere durch die Bereitstellung von Unterrichtsangeboten erreicht werden, die sich nicht an der durchschnittlichen Klassenleistung orientieren, sondern spezifisch für einzelne Schüler/innen oder für Schülergruppen mit einem sehr ähnlichen Leistungsstand konzipiert sind. Solche unterschiedlichen Lernangebote für verschiedene Schülergruppen innerhalb einer Klasse werden in der erziehungswissenschaftlichen Forschung als binnendifferenzierende Unterrichtsangebote bezeichnet (Roßbach & Wellenreuther, 2002).

Obwohl die Forderung nach Binnendifferenzierung auf bildungspolitischer Ebene und im wissenschaftlichen Diskurs immer wieder auftritt, deuten empirische Studien darauf hin, dass in Deutschland nur selten binnendifferenzierend unterrichtet wird. Für den Primarbereich zeigte sich in IGLU 2006 (Bos, Valtin, Voss, Hornberg & Lankes, 2007), dass nur etwa 20 % der befragten deutschen Lehrkräfte im Unterricht Aufgaben auf unterschiedlichem Anspruchsniveau einsetzen. Im Vergleich dazu berichteten knapp 80 % der englischen Lehrkräfte und etwa 60 % der schwedischen Lehrkräfte, auf das Leistungsniveau ihrer Schüler/innen abgestimmte Aufgaben zu verwenden (Lankes & Carstensen, 2007). Damit übereinstimmend deutet die Lehrerbefragung der Grundschulstudie VERA (in Rheinland-Pfalz und Bremen) darauf hin, dass Differenzierung vor allem in Form von Zusatzangeboten für leistungsstärkere, aber auch für leistungsschwächere Schüler/innen erfolgt, während Arbeit in homogenen Gruppen mit niveauangepassten Aufgaben innerhalb von Klassen seltener stattfindet (Schrader & Helmke, 2008; Helmke, 2014).

Für den Sekundarbereich existieren ähnliche Befunde. In PISA 2012 berichtete mehr als die Hälfte der Schüler/innen, dass im Mathematikunterricht *nie oder fast nie* Aufgaben auf unterschiedlichem Leistungsniveau angeboten werden (Schiepe-Tiska et al., 2013). Weitere 20 % der Schüler/innen gaben an, dass dies nur *in einigen Stunden* vorkomme. Eine Lehrerbefragung in PISA 2009 ergab ebenfalls, dass Differenzierung im Deutschunterricht selten stattfindet (Hertel, Hochweber, Steinert & Klieme, 2010). Ergebnisse der Lehrerbefragung im Rahmen der DESI-Studie deuten darauf hin, dass die häufigste Form der Differenzierung im Englischunterricht aus zusätzlichen Angeboten für leistungsstarke Schüler/innen besteht, während unterschiedliche Aufgaben für Einzelarbeitsphasen seltener vergeben werden (Helmke et al., 2008). Bezeichnend ist die Bilanz der Schulinspektion in Hessen, in der auf Basis von systematischen Unterrichts-



beobachtungen festgestellt wurde, dass differenzierende Unterrichtsangebote „in Form von qualitativ oder quantitativ unterschiedlichen Aufgaben und Aufträgen bzw. Lernmaterialien“ an keiner Schulform „gängige Praxis“ sind (Nieder & Frühauf, 2012, S. 26).

## *1.2 Mehr Binnendifferenzierung in leistungsheterogenen Klassen und bei guten Diagnostiker/innen?*

Empirische Forschung zu den Bedingungen für eine verstärkte Binnendifferenzierung ist bislang rar. Diskutiert werden insbesondere Kontextbedingungen wie das Ausmaß von Leistungsheterogenität in der Klasse sowie Überzeugungen und Kompetenzen der Lehrkräfte. So ist es plausibel, dass Lehrkräfte in sehr leistungsheterogenen Klassen stärker differenzieren, weil ein Unterricht, der sich am Durchschnittsschüler orientiert, besonders viele Schüler/innen der Klasse nicht erreicht. Warwas, Hertel und Labuhn (2011) argumentieren aber, dass solche kontextuellen Bedingungen allein die Unterrichtspraxis nicht erklären können, sondern dass auch professionelle Merkmale der Lehrkräfte in Betracht gezogen werden müssen. Die Autoren untersuchten den Zusammenhang zwischen Binnendifferenzierung und konstruktivistischen Überzeugungen, also der Auffassung, dass Schüler/innen die Konstrukteure des eigenen Wissens sind. Sie konnten zeigen, dass Lehrkräfte Aufgabenmaterial umso stärker an die Lernbedürfnisse der Schüler/innen anpassen, je stärker sie konstruktivistische Überzeugungen aufweisen (für ähnliche Befunde siehe auch Hofmann & Gottein, 2011). Lehrkräfte mit stärker ausgeprägten konstruktivistischen Überzeugungen variierten ihr Aufgabenmaterial insbesondere dann, wenn sie besonders leistungsheterogene Schülergruppen unterrichteten. Die Leistungsheterogenität allein schien hingegen nicht prädiktiv für den Einsatz von Binnendifferenzierung zu sein. Die Autoren weisen allerdings darauf hin, dass es sich bei ihren Befunden um Selbstberichtsdaten der Lehrkräfte handelt, die anhand von Videos oder externen Beobachtungen validiert werden sollten. Unberücksichtigt blieb auch, wie Lehrkräfte die Leistungsheterogenität in ihrer Klasse wahrnehmen. Möglicherweise unterrichten Lehrkräfte, die ihre Klasse als sehr leistungsheterogen wahrnehmen, stärker binnendifferenzierend als Lehrkräfte, die ihre Klasse als leistungshomogener erleben – auch wenn die Wahrnehmung nicht der objektiv messbaren Leistungsheterogenität der Klasse entspricht.

Als weitere wesentliche Voraussetzung für adaptiven und differenzierenden Unterricht werden diagnostische Kompetenzen von Lehrkräften genannt (z. B. Anders, Kunter, Brunner, Krauss & Baumert, 2010; Brunner, Anders, Hachfeld & Krauss, 2011; Helmke, 2014; Schrader, 1989). Diagnostische Kompetenzen bezeichnen die Fähigkeiten, leistungsrelevante Schülermerkmale sowie das Anspruchsniveau von Aufgaben angemessen einzuschätzen (Brunner et al., 2011). Sie werden als Schlüsselkompetenz für erfolgreiches Unterrichten angesehen, da sie erforderlich sind, um die Lernstände der Schüler/innen im Lernprozess fortlaufend festzustellen und Aufgaben bereitzustellen, die ein angemessenes Anforderungsniveau aufweisen. Entsprechend deuten empirische Studien darauf hin, dass hohe diagnostische Kompetenzen zu einer besseren Leis-

tungsentwicklung der Schüler/innen beitragen können (Anders et al., 2010; Helmke & Schrader, 1987; Karing, Pfof & Artelt, 2011). Anders et al. (2010) konnten für das Fach Mathematik zeigen, dass Lehrkräfte mit höheren diagnostischen Kompetenzen insofern adaptiver unterrichten, als sie in ihrem Unterricht Aufgaben mit einem höheren Aktivierungspotenzial einsetzen. Auch für die Bereitstellung differenzierender niveauangemessener Unterrichtsangebote erscheint die korrekte Einschätzung der Lernstände unabdingbar. Diese Annahme ist zwar plausibel, bisher fehlen aber empirische Belege dafür, dass ‚gute Diagnostiker/innen‘ ihren Unterricht tatsächlich stärker differenzierend gestalten.

### 1.3 Ziele dieser Arbeit und Hypothesen

Binnendifferenzierung wird als wesentliches Element einer adaptiven Unterrichtsgestaltung angesehen und kann eine Passung zwischen den heterogenen Lernvoraussetzungen von Schüler/innen innerhalb einer Lerngruppe einerseits und den Lernangeboten andererseits ermöglichen. Insbesondere an weiterführenden Schulen ist binnendifferenzierendes Lehrerhandeln bislang selten untersucht, obwohl empirische Studien zeigen konnten, dass auch innerhalb von Sekundarschulklassen und auch an Gymnasien substantielle Varianz im Leseverständnis und in mathematischen Kompetenzen vorliegt (Gröhlich et al., 2009). In dieser Studie stellen wir deshalb zunächst Ergebnisse zum Umfang binnendifferenzierenden Lehrerhandelns im Mathematik- und Deutschunterricht des 8. Jahrgangs an Gymnasien im Land Brandenburg vor. Wir fokussieren dabei auf den Einsatz leistungsdifferenzierender Aufgaben im Unterricht als eine wesentliche Möglichkeit der Differenzierung. Dafür berücksichtigen wir die Schülerperspektive auf den Unterricht. Unterrichtswahrnehmungen von Schüler/innen der Sekundarstufe gelten als zuverlässig und valide, insbesondere dann, wenn sie verhaltensnahe Beobachtungen des Unterrichts erfragen (Prenzel & Lankes, 2013), und können die Lehrerperspektive auf den Unterricht sinnvoll ergänzen.

Binnendifferenzierende Maßnahmen besitzen umso größere Relevanz, je heterogener die Leistungen innerhalb einer Klasse sind. Wir prüfen deshalb, ob der Umfang der Binnendifferenzierung zwischen Klassen und Lehrkräften variiert und ob ein Zusammenhang zwischen der von den Lehrkräften wahrgenommenen Leistungsheterogenität und der Binnendifferenzierung besteht. Unsere Hypothese ist dabei, dass Lehrkräfte in Klassen, die sie als leistungsheterogener wahrnehmen, in stärkerem Umfang leistungsdifferenzierende Aufgaben einsetzen als in Klassen, die sie als leistungshomogener wahrnehmen. Da weiterhin Diagnosekompetenzen der Lehrkräfte als wesentliche Voraussetzung für Binnendifferenzierung im Unterricht angesehen werden, untersuchen wir die Zusammenhänge zwischen der Diagnosegenauigkeit der Lehrkräfte und der Binnendifferenzierung. Wir nehmen an, dass Lehrkräfte, die das Leistungsniveau ihrer Schüler/innen adäquater einschätzen, in stärkerem Maße binnendifferenzierend unterrichten.

## 2. Methode

### 2.1 Stichprobe

Die Daten stammen aus einem größeren Forschungsprojekt zur Unterrichtsqualität, den Schulleistungen und dem sozio-emotionalen Befinden von Schüler/innen der Sekundarstufe I an Gymnasien (und einer Gesamtschule) im Land Brandenburg, die Klassen mit besonders leistungsstarken Schüler/innen führen. Die Studie wurde im Schuljahr 2012/13 durchgeführt und es nahmen 34 Schulen (42 % der staatlichen Gymnasien im Land Brandenburg) mit je zwei Klassen des 8. Jahrgangs teil, von denen eine aus besonders leistungsstarken Schüler/innen bestand. Die Teilnahme war für die Schulen verpflichtend. Die Stichprobe umfasste  $N = 1803$  Schüler/innen in 68 Klassen. Es nahmen 131 Lehrkräfte an der Untersuchung teil, von denen fünf eine Klasse in beiden Fächern oder ein Fach in zwei untersuchten Klassen unterrichteten. Die teilnehmenden Mathematiklehrkräfte verfügten über durchschnittlich 27 Jahre Berufserfahrung ( $SD = 7.1$ ) und 62 % von ihnen waren weiblich. Die teilnehmenden Deutschlehrkräfte besaßen im Mittel 22 Jahre Berufserfahrung ( $SD = 10.5$ ), 88 % von ihnen waren weiblich. Die Schüler/innen waren durchschnittlich 13.8 Jahre alt ( $SD = .49$ ), 53 % waren weiblich.

### 2.2 Instrumente

#### Schülerratings der Binnendifferenzierung

Die Schüler/innen schätzten die Binnendifferenzierung separat für den Mathematik- und Deutschunterricht anhand der Skala *Differenzierung* ein, die im Selbstevaluationsportal (SEP) des Instituts für Schulqualität der Länder Berlin und Brandenburg (ISQ) veröffentlicht ist (Gärtner, 2010a, 2010b). Die Skala bestand aus je vier Items für Mathematik und Deutsch („Mein Mathematiklehrer/meine Mathematiklehrerin gibt uns unterschiedliche Aufgaben, je nach unserem Können“, „Mein Mathematiklehrer/meine Mathematiklehrerin stellt Arbeitsgruppen nach unserem Können zusammen“, „Mein Mathematiklehrer/meine Mathematiklehrerin stellt den besseren Schülerinnen/Schülern schwierigere Aufgaben“, „Bei Klassenarbeiten können wir zwischen verschiedenen schwierigen Arbeiten auswählen“). Die Antwortskala war vierstufig mit 1 = „trifft überhaupt nicht zu“ bis 4 = „trifft voll und ganz zu“. Die Übereinstimmung der Ratings innerhalb einer Klasse prüften wir mit dem Index der mittleren Abweichung ( $AD_M$ ) (Lüdtke, Trautwein, Kunter & Baumert, 2006). Er wird berechnet, indem die mittlere Abweichung der Schülerratings vom Klassenmittelwert gebildet wird. Für das Fach Mathematik betrug der Index  $AD_M = .57$  und für das Fach Deutsch  $AD_M = .50$ . Im Schnitt wichen die individuellen Ratings demnach etwa einen halben Skalenpunkt vom Klassenmittelwert ab. Um die Reliabilität des mittleren Ratings der Klasse zu bestimmen, wurde die  $ICC(2)$  (Lüdtke et al., 2006) gebildet, die mit Cronbachs Alpha vergleichbar ist. Die Reliabilität der Schülerratings war zufriedenstellend und betrug für das Fach Mathematik  $ICC(2) = .88$  und für das Fach Deutsch  $ICC(2) = .75$ .

### **Mathematik- und Deutschleistungen**

Die Mathematik- und Deutschleistungen wurden im Rahmen der Vergleichsarbeiten in der Jahrgangsstufe 8 (VERA-8) erhoben. Sie prüfen, welche Kompetenzen die Schüler/innen zu diesem Zeitpunkt ihrer schulischen Laufbahn erworben haben. Die Entwicklung der Aufgaben stützte sich auf wissenschaftliche Kompetenzstufenmodelle und schloss eine umfangreiche Pilotierung und Normierung ein (Köller, 2008). Die Kompetenzen der Schüler/innen wurden separat für Mathematik und Deutsch auf einer Skala mit einem Mittelwert von 500 Punkten und einer Standardabweichung von 100 Punkten bestimmt. Anhand von Grenzwerten wurde die Leistung jedes/r Schüler/in anschließend einer von fünf Kompetenzstufen (von I = „unterer Mindeststandard“ bis V = „Maximalstandard“) zugeordnet (ISQ, 2012a).

### **Genauigkeit der Lehrereinschätzungen von Mathematik- und Deutschleistungen**

Die Mathematik- und Deutschlehrkräfte wurden gebeten, die Mathematik- bzw. Leseleistungen jedes/r Schüler/in ihrer Klasse anhand des Kompetenzstufenmodells einzuschätzen. Die Instruktion für die Mathematiklehrkräfte lautete: „Schätzen Sie bitte ein, welcher Kompetenzstufe (I bis V) die Mathematikleistungen jedes Schülers/jeder Schülerin Ihrer Klasse entsprechen“. Die Instruktion für die Deutschlehrkräfte war analog, jedoch auf die Leseleistungen bezogen. Dazu erhielten die Lehrkräfte eine schriftliche Erläuterung der Kompetenzstufen im Fach Mathematik und im Lesen, die wir aus einer Broschüre des ISQ (2012b) entnommen haben.

Wir berechneten drei verschiedene Indikatoren der Diagnosegenauigkeit: Die Fähigkeit der Lehrkräfte, die Schülerleistungen in der Klasse in eine Rangfolge zu bringen (Rangkomponente), operationalisierten wir über die Rangkorrelationen zwischen Lehrerurteilen und Testleistungen. Auch bei angemessener Einschätzung der Rangfolge ist jedoch eine deutliche Unter- oder Überschätzung des Leistungsniveaus möglich. Die Genauigkeit der Einschätzung des Leistungsniveaus der Schüler/innen (Niveauelemente) wurde über die mittlere Abweichung zwischen den Schülerleistungen und den Urteilen einer Lehrkraft berechnet. Zudem berechneten wir den prozentualen Anteil der Schüler/innen, die von jeder Lehrkraft korrekt eingeschätzt wurden. Dieser Anteil korrekt eingeschätzter Schüler/innen bildet ebenso wie die Niveauelemente vor allem die Fähigkeit ab, Schülerleistungen auf angemessenen Kompetenzniveaus zu verorten, statt, wie die Rangkomponente, Schülerleistungen in Relation zueinander zu beurteilen.

## **2.3 Datenanalyse**

Da unsere Daten eine hierarchische Struktur aufweisen, sind die Messungen nicht unabhängig voneinander, was jedoch eine Voraussetzung für OLS (ordinary least-square) Regression ist. Die Verletzung dieser Voraussetzung kann zur Unterschätzung der Standardfehler und zur Überschätzung der  $p$ -Werte führen. Aus diesem Grund führten wir hierarchische Regressionsanalysen durch, die berücksichtigen, dass die Messun-

gen nicht unabhängig voneinander sind. In diesen hierarchischen Regressionsanalysen modellierten wir die Schülerratings simultan auf Schülerebene und auf Klassenebene. Diese Analysen berechneten wir mit dem Statistikprogramm Mplus 6.1 (Muthén & Muthén, 1998–2010). Da unsere Fragestellungen Effekte auf Klassenebene fokussieren, berichten wir lediglich Ergebnisse auf Klassenebene.

Der Anteil der fehlenden Werte für die Schülerratings der Binnendifferenzierung, die Mathematik- und Deutschleistungen sowie die Lehrereinschätzungen betrug durchschnittlich 10.6% (Spannweite von 3.7 bis 18.2%). Bei den demografischen Angaben der Lehrkräfte fehlten 16% der Werte. Die fehlenden Werte wurden mit dem Ansatz der Chained Equations (MICE; van Buuren & Oudshoorn, 2000) im Statistikprogramm R 2.15.0 (R Development Core Team, 2012) multipel imputiert. Daten auf Schüler- und Klassenebene wurden separat imputiert. Für die Imputation der Schülerdaten wurden Klassenmittelwerte der Mathematik- und Deutschleistungen sowie des sozioökonomischen und kulturellen Kapitals als Prädiktoren einbezogen, um der geclusterten Struktur der Daten Rechnung zu tragen. Die fünf imputierten Datensätze wurden simultan in die Analysen einbezogen. Alle Ergebnisse basieren auf den imputierten Datensätzen.

**3. Ergebnisse**

*3.1 Binnendifferenzierung im Mathematik- und Deutschunterricht*

In Tabelle 1 sind die deskriptiven Statistiken für die auf Klassenebene aggregierten Schülerratings der Binnendifferenzierung dargestellt. Zusätzlich zu den Mittelwerten und Standardabweichungen ist der Range abgebildet. Das Ausmaß an Binnendifferenzierung wird von den Schüler/innen insgesamt als eher gering eingeschätzt, variiert jedoch zwischen den Klassen. Die Intraklassenkorrelationen für die Ratings der Binnendifferenzierung zeigen, dass zwischen Mathematiklehrkräften substanzielle Unterschiede in der Binnendifferenzierung ( $ICC = .21$ ) bestehen. Auch zwischen Deutschlehrkräften zeigen sich substanzielle Unterschiede in der Binnendifferenzierung ( $ICC = .10$ ).

Um zu prüfen, ob Mathematiklehrkräfte stärker differenzierend unterrichten als Deutschlehrkräfte, berechneten wir eine hierarchische Regressionsanalyse mit den Ratings der Binnendifferenzierung als abhängiger Variable und dem Unterrichtsfach (Deutsch = 0, Mathematik = 1) als unabhängiger Variable. Das Unterrichtsfach hatte

	<i>M</i>	<i>SD</i>	Range
Mathematik	1.88	0.40	1.26–3.26
Deutsch	1.62	0.25	1.14–2.30

Anmerkungen.  $N_{Klassen} = 68$  ( $N_{Schüler/innen} = 1803$ ).

Tab. 1: Mittelwerte, Standardabweichungen und Range der Binnendifferenzierung (Schülerratings auf Klassenebene aggregiert)

einen statistisch signifikanten Effekt auf die Schülerratings der Binnendifferenzierung ( $\beta = .26, p < .001$ ). Die Schüler/innen schätzten die Binnendifferenzierung im Fach Mathematik demnach durchschnittlich um etwa einen viertel Skalenpunkt höher ein als die Binnendifferenzierung in Deutsch.

3.2 *Wahrgenommene Leistungsheterogenität und Binnendifferenzierung*

In Tabelle 2 sind das mittlere, das minimale und das maximale Leistungsniveau der Klassen im Mathematik und Deutsch dargestellt. Außerdem haben wir zwei Indikatoren der Leistungsheterogenität der Klassen gebildet, indem wir zunächst die Streuung (Standardabweichung und Range) innerhalb jeder Klasse bestimmt und anschließend über alle Klassen gemittelt haben. Das mittlere Leistungsniveau liegt in Mathematik bei  $M = 3.77$  und in Deutsch bei  $M = 4.25$ . Die mittleren Lehrerurteile liegen in beiden Fächern deutlich unterhalb des tatsächlichen mittleren Leistungsniveaus (Tab. 2). Die Leistungsspanne zwischen dem/der leistungsschwächsten und dem/der leistungstärksten Schüler/in innerhalb einer Klasse liegt bei durchschnittlich knapp drei Kompetenzstufen in Mathematik und etwa zwei Kompetenzstufen in Deutsch. Die mittlere Standardabweichung und das mittlere Range der Lehrerurteile sind in Mathematik und Deutsch marginal höher als für die Leistungswerte. Die Lehrkräfte überschätzten folglich die Leistungsheterogenität ihrer Schüler/innen leicht.

Unsere erste Hypothese war, dass Lehrkräfte stärker differenzieren, wenn sie ihre Klasse als leistungsheterogener wahrnehmen. Um diese Hypothese zu prüfen, führten wir separat für beide Fächer hierarchische Regressionsanalysen durch, in denen wir die Schülerratings der Binnendifferenzierung durch die wahrgenommene Leistungshetero-

	<i>M (SD)</i>	<i>Min</i>	<i>Max</i>	Mittlere Hetero- genität (SD)	Mittlere Hetero- genität (Range)
				innerhalb der Schulklassen <sup>1</sup>	
Leistungen in VERA					
Mathematik	3.77 (0.58)	2.53	4.91	0.79	2.72
Deutsch	4.25 (0.53)	2.58	4.97	0.67	2.17
Lehrerurteile					
Mathematik	3.19 (0.57)	1.87	4.78	0.82	2.85
Deutsch	3.26 (0.52)	2.23	4.38	0.75	2.55

Anmerkungen. <sup>1</sup> Standardabweichung/Range der Leistungen innerhalb der Klassen bestimmt und über alle Klassen gemittelt.  
 $N_{Klassen} = 68$ .

Tab. 2: *Leistungsniveau und Leistungsheterogenität der Schülerleistungen und Lehrerurteile in Mathematik und Deutsch*

genität (mittlere Standardabweichung der Lehrerurteile innerhalb der Klassen) vorher-sagten. Die wahrgenommene Leistungsheterogenität lieferte weder für das Fach Ma-thematik noch für Deutsch einen statistisch signifikanten Beitrag zur Aufklärung der Varianz in der Binnendifferenzierung auf Klassenebene (Tab. 4 und 5, Modell 1). Kon-trollanalysen zeigten, dass sich die Binnendifferenzierung auch durch die in VERA ge-messene Leistungsheterogenität (mittlere Standardabweichung der Testleistungen in-nerhalb der Klassen) nicht signifikant vorhersagen ließ ( $p > .05$ , nicht in Tabelle). Wie stark Lehrkräfte leistungsdifferenzierende Aufgaben einsetzen, war folglich nicht mit der von ihnen wahrgenommenen oder der tatsächlichen Leistungsheterogenität der Klasse assoziiert.

### 3.3 Binnendifferenzierung und Diagnosegenauigkeit

In Tabelle 3 sind die deskriptiven Statistiken für die drei Indikatoren der Diagnose-genauigkeit dargestellt. Im Mittel konnten Mathematiklehrkräfte den Kompetenzstand von 37% ihrer Schüler/innen der korrekten Kompetenzstufe zuordnen; Deutschlehr-kräften gelang dies bei durchschnittlich 24% ihrer Schüler/innen. Lehrkräfte unter-schätzten das mittlere Leistungsniveau in Mathematik um durchschnittlich etwas mehr als eine halbe Kompetenzstufe und in Deutsch um eine ganze Kompetenzstufe (Ni-veaukomponente). Die Fähigkeiten der Lehrkräfte, eine korrekte Rangfolge der Schü-lerleistungen zu bilden, bestimmt über die Rangkorrelationen zwischen Lehrerurteilen und Testleistungen, lagen im Fach Mathematik bei  $r = 0.38$  und in Deutsch bei  $r = 0.21$ .

Unsere zweite Hypothese postulierte, dass Lehrkräfte mit höheren diagnostischen Fähigkeiten stärker binnendifferenzierend unterrichten. Um diese Hypothese zu un-tersuchen, berechneten wir separat für Deutsch und Mathematik hierarchische Regres-sionsanalysen, in denen die Schülerratings der Binnendifferenzierung durch die Dia-gnosegenauigkeit vorhergesagt wurden. Als Indikatoren der Diagnosegenauigkeit berücksichtigten wir die (vom Betrag her invertierte) Niveauekomponente, den Anteil korrekt eingeschätzter Schüler/innen und die Rangkomponente. Positive Effekte indi-zieren folglich einen positiven Zusammenhang zwischen Diagnosegenauigkeit und Bin-

	Mathematik				Deutsch			
	<i>M</i>	( <i>SD</i> )	<i>Min</i>	<i>Max</i>	<i>M</i>	( <i>SD</i> )	<i>Min</i>	<i>Max</i>
Anteil korrekt (%)	36.5	(15.2)	3.9	71.8	23.6	(14.7)	2.2	62.5
Niveauekomponente	-0.58	(0.48)	-1.68	0.43	-0.98	(0.58)	-2.17	0.27
Rangkomponente	0.38	(0.24)	-0.60	0.74	0.21	(0.26)	-0.45	0.73

Anmerkungen.  $N_{Klassen} = 68$ .

Tab. 3: Mittelwerte, Standardabweichungen und Spannweiten der Indikatoren der Diagnose-genauigkeit



	Modell 1		Modell 2		Modell 3		Modell 4	
	B	SE	B	SE	B	SE	B	SE
<i>Klassenebene</i>								
Geschlecht <sup>1</sup>	-.15	.13	-.12	.13	-.15	.13	-.15	.13
Berufserfahrung	.02	.14	.07	.14	.04	.14	.01	.14
Wahrgenommene Heterogenität <sup>2</sup>	-.02	.17						
Diagnosegenauigkeit (Niveau) <sup>3</sup>			.28*	.11				
Diagnosegenauigkeit (Anteil)					.26*	.12		
Diagnosegenauigkeit (Rang)							.04	.12
R <sup>2</sup> zwischen den Klassen	.03		.10		.10		.03	

Anmerkungen. Standardisierte Ergebnisse.  $N_{\text{Schüler/innen}} = 1803$ ,  $N_{\text{Klassen}} = 68$ . <sup>1</sup> Referenz: weiblich. <sup>2</sup> Standardabweichung der Lehrerurteile. <sup>3</sup> (vom Betrag her invertierte) Niveauelemente. \*  $p < .05$ .

Tab. 4: Hierarchisch lineare Vorhersage der Binnendifferenzierung in Mathematik durch die wahrgenommene Leistungsheterogenität in der Klasse und die Diagnosegenauigkeit der Lehrkraft

	Modell 1		Modell 2		Modell 3		Modell 4	
	B	SE	B	SE	B	SE	B	SE
<i>Klassenebene</i>								
Geschlecht <sup>1</sup>	-.06	.15	-.04	.14	-.03	.14	-.04	.15
Berufserfahrung	-.18	.18	-.11	.18	-.11	.17	-.15	.18
Wahrgenommene Heterogenität <sup>2</sup>	-.17	.15						
Diagnosegenauigkeit (Niveau) <sup>3</sup>			.16	.12				
Diagnosegenauigkeit (Anteil)					.20	.13		
Diagnosegenauigkeit (Rang)							.00	.15
R <sup>2</sup> zwischen den Klassen	.06		.05		.07		.03	

Anmerkungen. Standardisierte Ergebnisse.  $N_{\text{Schüler/innen}} = 1803$ ,  $N_{\text{Klassen}} = 68$ . <sup>1</sup> Referenz: weiblich. <sup>2</sup> Standardabweichung der Lehrerurteile. <sup>3</sup> (vom Betrag her invertierte) Niveauelemente.

Tab. 5: Hierarchisch lineare Vorhersage der Binnendifferenzierung in Deutsch durch die wahrgenommene Leistungsheterogenität in der Klasse und die Diagnosegenauigkeit der Lehrkraft



nendifferenzierung. In den Tabellen 4 und 5 sind die standardisierten Ergebnisse dargestellt, die mithilfe von Mplus berechnet wurden, indem jedes Regressionsgewicht an der Standardabweichung der unabhängigen und der abhängigen Variable standardisiert wird. Im Fach Mathematik fanden sich positive Effekte der Diagnosegenauigkeit auf die Binnendifferenzierung (Tab. 4, Modelle 2 und 3). Lehrkräfte, die das Leistungsniveau der Klasse adäquater einschätzten bzw. einen größeren Anteil der Schüler/innen ihrer Klasse korrekt auf Kompetenzstufen verorten konnten, führten aus Schülerperspektive stärker differenzierenden Unterricht durch. In Deutsch fanden sich keine analogen Effekte (Tab. 5, Modelle 2 und 3, alle  $p > .05$ ). Die Rangkomponente korrespondierte hingegen weder im Fach Mathematik (Tab. 4, Modell 4) noch im Fach Deutsch (Tab. 5, Modell 4) mit der Binnendifferenzierung im Unterricht.

#### 4. Diskussion

Der Umgang mit Heterogenität wird als eine der zentralen Herausforderungen des Bildungssystems angesehen. Auch an weiterführenden Schulen stellt sich Lehrkräften alltäglich die Frage, wie sie der Leistungsheterogenität begegnen können, die sie innerhalb ihrer Schulklasse vorfinden. Damit möglichst alle Schüler/innen ihrem Entwicklungsstand angemessen gefördert werden können, sollen Lehrkräfte binnendifferenzierende Elemente in ihren Unterricht integrieren. Die vorliegende Studie untersuchte, in welchem Umfang Gymnasiallehrkräfte in Brandenburg im Mathematik- und Deutschunterricht der 8. Jahrgangsstufe das Aufgabenmaterial auf die unterschiedlichen Kompetenzen ihrer Schüler/innen abstimmen. Insbesondere prüften wir, ob Lehrkräfte stärker differenzieren, wenn sie erstens eine größere Leistungsheterogenität in ihrer Klasse wahrnehmen und zweitens über höhere Diagnosekompetenzen verfügen.

In unserer Untersuchung zeigte sich, dass Lehrkräfte aus Schülersicht nur in geringem Umfang leistungsdifferenzierende Aufgaben einsetzen. Die Schülereinschätzungen in unserer Stichprobe stimmen folglich mit den Ergebnissen aus Schulinspektionen (z.B. Nieder & Frühauf, 2012) und den Schülerbefragungen zum Mathematikunterricht in PISA (Schiepe-Tiska et al., 2013) überein. Unsere Ergebnisse deuten außerdem darauf hin, dass leistungsdifferenzierende Aufgaben im Mathematikunterricht in stärkerem Ausmaß eingesetzt werden als im Deutschunterricht. Eine mögliche Erklärung für die geringere Binnendifferenzierung im Deutschunterricht könnten die vielfältigen Lösungsmöglichkeiten sein, die im Deutschunterricht für ein und dieselbe Aufgabe bestehen können. Entsprechend ihrer individuellen Kompetenzen können Schüler/innen beispielsweise Texte auf ganz unterschiedlichem Niveau verfassen, sodass sich Schüler/innen im Deutschunterricht der Sekundarstufe möglicherweise seltener deutlich unterfordert oder überfordert fühlen.

Die von uns gefundene leichte Überschätzung der Leistungsheterogenität ist im Einklang mit Ergebnissen aus anderen Studien zur Diagnosegenauigkeit, die angemessene Einschätzungen (Brunner et al., 2011) oder Überschätzungen der Leistungsheterogenität berichteten (Leucht, Tiffin Richards, Vock, Pant & Köller, 2012). Entgegen unserer

Hypothese fanden wir allerdings keine Zusammenhänge zwischen der von den Lehrkräften wahrgenommenen oder der tatsächlichen Leistungsheterogenität und der Binnendifferenzierung. Dies erweitert die Befunde von Warwas et al. (2011), die keine Zusammenhänge zwischen der Leistungsheterogenität in Grundschulklassen und der von den Lehrkräften selbsteingeschätzten Binnendifferenzierung im Mathematik- und Deutschunterricht fanden. Heterogenität wahrzunehmen scheint folglich keine hinreichende Voraussetzung für differenzierendes Unterrichtshandeln. Für die Bereitstellung niveaueingepasster Aufgaben scheinen weitere Kompetenzen erforderlich zu sein.

Die über die Rangkomponente gemessene Diagnosegenauigkeit war in der vorliegenden Untersuchung eher niedrig (Südkamp, Kaiser & Möller, 2012), aber vergleichbar mit der Genauigkeit, die in einer ähnlich konzipierten Studie berichtet wurde, in der Lehrkräfte die Schülerleistungen auf den Kompetenzstufen des Gemeinsamen Europäischen Referenzrahmens verorten sollten (Leucht et al., 2012). Die Unterschätzung des Leistungsniveaus in unserer Studie steht im Gegensatz zu Befunden anderer Studien, die mehrheitlich auf eine Überschätzung des Leistungsniveaus hindeuten (zusammenfassend van Ophuysen, 2010). Vermutlich ist die von uns gefundene Unterschätzung darauf zurückzuführen, dass die Schülerstichprobe recht leistungsstark war. Denn eine zumindest tendenzielle Unterschätzung des Leistungsniveaus leistungsstarker Schülergruppen wurde auch in anderen Studien bereits nachgewiesen (Südkamp, Möller & Pohlmann, 2008; für ähnliche Ergebnisse siehe auch Brunner et al., 2011). Die insgesamt eher niedrige Diagnosegenauigkeit könnte auch ein Hinweis darauf sein, dass die Lehrkräfte das Kompetenzstufenmodell im schulischen Alltag bisher kaum verwenden und sich entsprechend noch wenig vertraut damit fühlen (Leucht et al., 2012). Unsere Hypothese, dass Lehrkräfte mit höheren diagnostischen Kompetenzen stärker differenzierend unterrichten, ließ sich teilweise bestätigen. Wir fanden, dass Mathematiklehrkräfte, die das Leistungsniveau der Schüler/innen ihrer Klasse adäquater einschätzen oder einen höheren Anteil der Schüler/innen korrekt auf Kompetenzstufen verorten können, eher leistungsdifferenzierende Aufgaben im Unterricht einsetzen als Lehrkräfte, die weniger gute Diagnostiker/innen sind. Dieses Ergebnis bestätigt einerseits die vielfach aufgeführte, aber kaum empirisch überprüfte Annahme, dass höhere diagnostische Fähigkeiten mit einer adaptiveren Unterrichtsgestaltung einhergehen, und spricht andererseits für die Validität der erhobenen Schülerratings. Die Zusammenhänge zeigten sich für das Fach Deutsch allerdings nicht. Möglicherweise sehen Deutschlehrkräfte, auch wenn sie über gute diagnostische Fähigkeiten verfügen, weniger Notwendigkeit für aufgabenbezogene Differenzierung, da Aufgaben im Deutschunterricht häufig vielfältige Lösungsmöglichkeiten erlauben, sodass Schüler/innen sie entsprechend ihren Kompetenzständen bearbeiten können. Zudem war die Fähigkeit, Schülerleistungen innerhalb der Klasse in eine Rangreihe zu bringen, nicht prädiktiv für den Umfang der Binnendifferenzierung in beiden hier untersuchten Schulfächern. Wie lässt sich dieses Ergebnis erklären? Auch wenn Lehrkräfte gut in der Lage sind, klasseninterne Reihungen vorzunehmen, besteht die Gefahr der systematischen Über- oder Unterschätzung des Leistungsniveaus ihrer Schüler/innen. In empirischen Studien zeigen die beiden Fähigkeiten entsprechend nur moderate Zusammenhänge (z. B. Schrader, Helmke, Hosenfeld, Halt

& Hochweber, 2006). Um leistungsangemessene Unterrichtsangebote bereitstellen zu können, ist vermutlich insbesondere die Einschätzung der Kompetenzstände der Schüler/innen über die klasseninterne Norm hinaus von Bedeutung.

#### 4.1 *Pädagogische Implikationen*

Aufgrund des querschnittlichen Designs unserer Untersuchung müssen die Ergebnisse vorsichtig interpretiert werden. Dennoch möchten wir mögliche Implikationen für die Lehrerprofessionalisierung aus unserer Studie ableiten. Da sich Lehrkräfte bei der Beurteilung von Schülerleistungen häufig an einer klasseninternen Norm orientieren, sind sie meist recht gut in der Lage, klasseninterne Beurteilungen vorzunehmen. Unsere Befunde unterstreichen vor dem Hintergrund der Annahme Helmkes (2014), der diagnostische Kompetenzen als Voraussetzung der Binnendifferenzierung sieht, wie wichtig es ist, Lehrkräften das Einschätzen von Schülerleistungen über den Bezugsrahmen ihrer eigenen Klasse hinaus nahezubringen. Hierfür sind in den letzten Jahren von der Bildungspolitik mit der Einführung der Bildungsstandards und der Vergleichsarbeiten (VERA) bereits erste Grundsteine gelegt worden. Die Frage ist aber, inwieweit diese Lerngelegenheiten von Lehrkräften auch tatsächlich genutzt werden bzw. genutzt werden können und ob hier nicht weitere Unterstützungssysteme benötigt würden. Denn Lehrkräfte verfügen häufig nicht über hinreichende statistische Kenntnisse, um komplexe Datenauswertungen zu verstehen (Hellrung & Hartig, 2013). Zudem nutzen sie Rückmeldungen bspw. im Rahmen von VERA insbesondere dann, wenn sie diese als Instrumente der Unterrichtsentwicklung wahrnehmen (Richter, Böhme, Becker, Pant & Stanat, 2014). Entsprechend scheint es von besonderer Relevanz, den Lehrkräften diese Funktion von Rückmeldungen verstärkt nahezubringen.

#### 4.2 *Limitationen*

Die vorliegende Untersuchung liefert Hinweise auf die empirischen Zusammenhänge zwischen leistungsdifferenzierendem Unterrichtshandeln und der Diagnosegenauigkeit der Lehrkräfte, weist jedoch auch Grenzen auf. Erstens konzentrierte sich die Studie auf den Gymnasialunterricht, da insbesondere an weiterführenden Schulen Bedingungen differenzierenden Unterrichtshandelns bislang kaum empirisch untersucht sind. Die Ergebnisse sind jedoch nicht ohne Weiteres auf andere Schulformen generalisierbar, denn Lehrkräfte an verschiedenen Schulformen begegnen der Leistungsheterogenität ihrer Schülerschaft mit tendenziell unterschiedlichen Formen differenzierenden Unterrichts (Helmke et al., 2008). Zweitens war die Untersuchung querschnittlich angelegt und erlaubt deshalb keine Aussagen über die Wirkrichtung der Zusammenhänge. In der Forschungsliteratur wird die Diagnosegenauigkeit häufig als Voraussetzung dafür angesehen, dass Lehrkräfte das Unterrichtsniveau an die Fähigkeiten ihrer Schüler/innen anpassen können (z. B. Helmke, 2014). Möglich wäre allerdings auch, dass Lehrkräfte, die

häufiger differenzierend unterrichten, die Fähigkeiten ihrer Schüler/innen besser einzuschätzen lernen. Auch eine wechselseitige Beeinflussung ist denkbar. Drittens wird das Ausmaß der Binnendifferenzierung ausschließlich aus Sicht der Schüler/innen erfasst. Die Schülerperspektive auf den Unterricht gilt als zuverlässig und valide, insbesondere, wenn verhaltensnahe Unterrichtsaspekte eingeschätzt werden sollen (Prenzel & Lankes, 2013). Dafür spricht auch die hohe Einigkeit der Schüler/innen einer Klasse in unserer Studie. Dennoch kann es vorkommen, dass Schüler/innen differenzierende Unterrichtsangebote nicht als solche wahrnehmen, beispielsweise, wenn niveauangepasste Aufgaben von den Lehrkräften diskret vergeben und nicht explizit gemacht werden (Gärtner, 2010a). Zudem weisen Schülereinschätzungen perspektivenspezifische Varianz auf (Fauth, Decristan, Rieser, Klieme & Büttner, 2014). Beispielsweise wird bei der Einschätzung, ob die Lehrkraft Aufgaben je nach dem Können der Schüler/innen verteilt, indirekt erfragt, inwiefern ein/e Schüler/in die ihm/ihr zugeteilten Aufgaben als leistungsangemessen empfindet. Weiterhin ist einschränkend anzumerken, dass gerade bei dem genannten Item möglicherweise eine Konfundierung mit der diagnostischen Kompetenz vorliegt. Zudem zielen die von uns verwendeten Items eher auf eine Form der Differenzierung, die Heymann (2010) als geschlossene Differenzierung versteht, bei der die Lehrkräfte ihren Schüler/innen niveauangemessene Lernangebote zuweisen, möglichst auf der Grundlage einer Diagnose. Für den Einsatz offener Differenzierungsformen, bei denen Schüler/innen ihre Aufgaben frei wählen, könnte die angemessene Einschätzung der Leistungsstreuung hinreichend sein, um Aufgaben bereitzustellen zu können, die das gesamte Leistungsspektrum bedienen können (van Ophuysen, 2010). Schließlich haben wir die Diagnosegenauigkeit der Lehrkräfte auf Grundlage eines globalen Urteils über Schülerkompetenzen in Mathematik und im Lesen bestimmt. Lorenz und Artelt (2009) konnten zeigen, dass Lehrkräfte verschiedene Teilkompetenzen innerhalb derselben Domäne (Wortschatz und Textverstehen) ähnlich genau beurteilen können. Jedoch ist zu beachten, dass Lehrkräfte für die Anpassung von Lernmaterialien an die individuellen Leistungsstände der Schüler/innen insbesondere im Fach Mathematik bereichsspezifisches Vorwissen beurteilen müssen (Helmke, 2014), das in verschiedenen mathematischen Stoffgebieten unterschiedlich sein kann (Brunner, 2006).

### 4.3 Fazit

Vor dem Hintergrund der aktuellen Heterogenitätsdebatte liefert die vorliegende Studie wichtige empirische Hinweise darauf, inwiefern lehrerseitige Kompetenzen mit differenzierendem Unterrichtshandeln verknüpft sind. Da für die Realisierung von adaptivem Unterricht auch didaktische Kompetenzen eine wesentliche Rolle spielen (Beck et al., 2008) und sich diagnostische Kompetenzen vor allem in Kombination mit spezifischen didaktischen Kompetenzen als lernförderlich erwiesen haben (Helmke & Schrader, 1987), sollten zukünftige Untersuchungen das Zusammenwirken diagnostischer und didaktischer Kompetenzen in den Blick nehmen. Hierfür könnten Interventionsstudien geeignet sein, in denen für differenzierendes Unterrichtshandeln notwendige didakti-

sche Kompetenzen gestärkt und kriteriale Diagnosekompetenzen, beispielsweise durch die intensive Auseinandersetzung mit externen Standards oder Referenzrahmen, geschult werden. Empirische Untersuchungen sollten dabei auch in den Blick nehmen, unter welchen Bedingungen sich verschiedene Formen von Binnendifferenzierung günstig auf die Entwicklung von Schulleistungen und sozioemotionalen Kompetenzen auswirken. Eine Metaanalyse von über 50 internationalen Studien deutete zwar darauf hin, dass binnendifferenzierender Unterricht im Vergleich zu nicht-differenzierendem Jahrgangsunterricht einen leicht förderlichen Effekt auf die Leistungsentwicklung haben kann (Lou et al., 1996). Einleuchtend ist aber, dass bestimmte Rahmenbedingungen und bestimmte Kompetenzen aufseiten der Lehrkräfte erforderlich sind, damit unerwünschte Nebenwirkungen ausbleiben (Gruehn, 2000). Zur Klärung der Zusammenhänge sind insbesondere Trainingsstudien vielversprechend (Beck et al., 2008).

## Literatur

- Anders, Y., Kunter, M., Brunner, M., Krauss, S., & Baumert, J. (2010). Diagnostische Fähigkeiten von Mathematiklehrkräften und ihre Auswirkungen auf die Leistungen ihrer Schülerinnen und Schüler. *Psychologie in Erziehung und Unterricht*, 57(3), 175–193.
- Beck, E., Baer, M., Guldemann, T., Bischoff, S., Brühwiler, Ch., Müller, P., Niedermann, R., Rogalla, M., & Vogt, F. (2008). *Adaptive Lehrkompetenz. Analyse von Struktur, Veränderbarkeit und Wirkung handlungssteuernden Lehrerwissens*. Münster: Waxmann.
- Bos, W., Valtin, R., Voss, A., Hornberg, S., & Lankes, E.-M. (2007). Konzepte der Lesekompetenz in IGLU 2006. In W. Bos, S. Hornberg, K.-H. Arnold, G. Faust, L. Fried, E.-M. Lankes, K. Schwippert & R. Valtin (Hrsg.), *IGLU 2006. Lesekompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich* (S. 81–107). Münster: Waxmann.
- Brunner, M. (2006). *Mathematische Schülerleistung: Struktur, Schulformunterschiede und Validität* (Dissertation). Berlin: Humboldt-Universität zu Berlin. <http://edoc.hu-berlin.de/dissertationen/brunner-martin-2006-02-08/HTML/front.html> [16.02.2015].
- Brunner, M., Anders, Y., Hachfeld, A., & Krauss, S. (2011). Diagnostische Fähigkeiten von Mathematiklehrkräften. In M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Krauss & M. Neubrand (Hrsg.), *Professionelle Kompetenz von Lehrkräften. Ergebnisse des Forschungsprogramms COACTIV* (S. 215–234). Münster: Waxmann.
- Ditton, H. (2000). Qualitätskontrolle und Qualitätssicherung in Schule und Unterricht. Ein Überblick zum Stand der empirischen Forschung. *Zeitschrift für Pädagogik*, 41. Beiheft, 73–92.
- Ditton, H. (2002). Lehrkräfte und Unterricht aus Schülersicht. Ergebnisse einer Untersuchung im Fach Mathematik. *Zeitschrift für Pädagogik*, 48(2), 262–286.
- Fauth, B., Decristan, J., Rieser, S., Klieme, E., & Büttner, G. (2014). Grundschulunterricht aus Schüler-, Lehrer- und Beobachterperspektive: Zusammenhänge und Vorhersage von Lernerfolg. *Zeitschrift für Pädagogik*, 28(3), 127–137.
- Gärtner, H. (2010a). Wie Schülerinnen und Schüler ihre Lernumwelt wahrnehmen: Ein Vergleich verschiedener Maße zur Übereinstimmung von Schülerwahrnehmungen. *Zeitschrift für Pädagogische Psychologie*, 24(2), 111–122.
- Gärtner, H. (2010b). Das ISQ-Selbstevaluationsportal. Konzeption eines Online-Angebots, um die Selbstevaluation in Schule und Unterricht zu unterstützen. *Die Deutsche Schule*, 102(2), 163–175.
- Gröhlisch, C., Scharenberg, K., & Bos, W. (2009). Wirkt sich Leistungsheterogenität in Schulklassen auf den individuellen Lernerfolg in der Sekundarstufe aus? *Journal for Educational Research Online*, 1(1), 86–105.

- Gruehn, S. (2000). *Unterricht und schulisches Lernen. Schüler als Quellen der Unterrichtsbeschreibung*. Münster: Waxmann.
- Hellrung, K., & Hartig, J. (2013). Understanding and using feedback. A review of empirical studies concerning feedback from external evaluations to teachers. *Educational Research Review*, 9, 174–190.
- Helmke, A. (2014). *Unterrichtsqualität und Lehrerprofessionalität: Diagnose, Evaluation und Verbesserung des Unterrichts* (5. Aufl.). Seelze: Klett-Kallmeyer.
- Helmke, T., Helmke, A., Schrader, F. W., Wagner, W., Nold, G., & Schröder, K. (2008). Alltagspraxis des Englischunterrichts. In DESI-Konsortium (Hrsg.), *Unterricht und Kompetenzerwerb in Deutsch und Englisch. Ergebnisse der DESI-Studie* (S. 371–381). Weinheim/Basel: Beltz.
- Helmke, A., & Schrader, F. W. (1987). Interactional effects of instructional quality and teacher judgement accuracy on achievement. *Teaching and Teacher Education*, 3(2), 91–98.
- Hertel, S., Hochweber, J., Steinert, B., & Klieme, E. (2010). Schulische Rahmenbedingungen und Lerngelegenheiten im Deutschunterricht. In E. Klieme, C. Artelt, J. Hartig, N. Jude, O. Köller, M. Prenzel, W. Schneider & P. Stanat (Hrsg.), *PISA 2009. Bilanz nach einem Jahrzehnt* (S. 113–149). Münster: Waxmann.
- Heymann, H. W. (2010). Binnendifferenzierung – eine Utopie? Pädagogischer Anspruch, didaktisches Handwerk, Realisierungschancen. *Pädagogik*, 62(11), 6–11.
- Hofmann, F., & Gottein, H.-P. (2011). Zusammenhänge zwischen Persönlichkeitsmerkmalen und der Bereitschaft zur Individualisierung im Unterricht. *Zeitschrift für Bildungsforschung*, 1(1), 55–67.
- Hohn, K., Schiepe-Tiska, A., Sälzer, C., & Artelt, C. (2013). Lesekompetenz in PISA 2012: Veränderungen und Perspektiven. In M. Prenzel, C. Sälzer, E. Klieme & O. Köller (Hrsg.), *PISA 2012. Fortschritte und Herausforderungen in Deutschland* (S. 217–244). Münster: Waxmann.
- Institut für Schulqualität der Länder Berlin und Brandenburg (ISQ) (2012a). *Vergleichsarbeiten der Jahrgangsstufe 8 in Brandenburg als Beitrag zur Schul- und Unterrichtsentwicklung* (Online-Publikation). [http://www.isq-bb.de/uploads/media/Broschuere\\_BRB\\_VERA\\_8\\_2012-11-15.pdf](http://www.isq-bb.de/uploads/media/Broschuere_BRB_VERA_8_2012-11-15.pdf) [27. 11. 2013].
- Institut für Schulqualität der Länder Berlin und Brandenburg (ISQ) (2012b). *Vergleichsarbeiten in der Jahrgangsstufe 8 im Schuljahr 2011/2012. Länderbericht Brandenburg*. [http://www.isq-bb.de/uploads/media/Bericht\\_BRB\\_2012\\_12\\_10\\_yy\\_Ansicht\\_Druckfassung\\_FM.pdf](http://www.isq-bb.de/uploads/media/Bericht_BRB_2012_12_10_yy_Ansicht_Druckfassung_FM.pdf) [11. 02. 2014].
- Karing, C., Pfost, M., & Artelt, C. (2011). Hängt die diagnostische Kompetenz von Sekundarstufenlehrkräften mit der Entwicklung der Lesekompetenz und der mathematischen Kompetenz ihrer Schülerinnen und Schüler zusammen? *Journal für Bildungsforschung Online*, 3(2), 119–147.
- Köller, O. (2008). Bildungsstandards in einem Gesamtsystem der Qualitätssicherung im allgemein bildenden Schulsystem Deutschlands. *Zeitschrift für Pädagogik*, 53. Beiheft, 59–75.
- Lankes, E.-M., & Carstensen, C. H. (2007). Der Leseunterricht aus der Sicht der Lehrkräfte. In W. Bos, S. Hornberg, K.-H. Arnold, G. Faust, L. Fried, E.-M. Lankes, K. Schwippert & R. Valtin (Hrsg.), *IGLU 2006. Lesekompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich* (S. 161–193). Münster: Waxmann.
- Leucht, M., Tiffin Richards, S., Vock, M., Pant, H. A., & Köller, O. (2012). Diagnostische Kompetenz von Englischlehrkräften bei der Bewertung von Schülerleistungen mit Hilfe des Gemeinsamen Europäischen Referenzrahmens für Sprachen. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 44(4), 163–177.
- Lorenz, C., & Artelt, C. (2009). Fachspezifität und Stabilität diagnostischer Kompetenz von Grundschullehrkräften in den Fächern Deutsch und Mathematik. *Zeitschrift für Pädagogische Psychologie*, 23(34), 211–222.



- Lou, Y., Abrami, P. C., Spence, J. C., Poulsen, C., Chambers, B., & d'Apollonia, S. (1996). Within-class grouping: a meta-analysis. *American Educational Research*, 66(4), 423–458.
- Lüdtke, O., Trautwein, U., Kunter, M., & Baumert, J. (2006). Reliability and agreement of student ratings of the classroom environment: A reanalysis of TIMSS data. *Learning Environments Research*, 9(3), 215–230.
- Muthén, B. O., & Muthén, L. K. (1998–2010). *Mplus Version 6.1. Base program and multilevel add-on* [computer program]. Los Angeles: Muthén & Muthén.
- Nieder, T., & Frühauf, S. (2012). *Bilanzbericht der Schulinspektion. Ergebnisse der externen Evaluation an allen hessischen Schulen*. [http://www.deqa-vet.de:8001/\\_media/PDF\\_allgemein/Bilanzbericht\\_Hessen\\_2012.pdf](http://www.deqa-vet.de:8001/_media/PDF_allgemein/Bilanzbericht_Hessen_2012.pdf) [30.06.2014].
- Prenzel, M., & Lankes, E.-M. (2013). Was können Schülerinnen und Schüler über ihren Unterricht sagen? Ein Blick in die Schülerfragebögen von internationalen Vergleichsstudien. In N. McElvany & H. Holtappels (Hrsg.), *Empirische Bildungsforschung. Theorien, Methoden, Befunde und Perspektiven* (S. 93–107). Münster: Waxmann.
- R Development Core Team (2012). *R: A language and environment for statistical computing*. Wien: R Foundation for Statistical Computing.
- Richter, D., Böhme, K., Becker, M., Pant, H. A., & Stanat, P. (2014). Überzeugungen von Lehrkräften zu den Funktionen von Vergleichsarbeiten. *Zeitschrift für Pädagogik*, 60(2), 225–243.
- Roßbach, H.-G., & Wellenreuther, M. (2002). Empirische Forschungen zur Wirksamkeit von Methoden der Leistungsdifferenzierung in der Grundschule. In F. Heinzel & A. Prengel (Hrsg.), *Heterogenität, Integration und Differenzierung in der Primarstufe* (Bd. 6, S. 44–57). Opladen: Leske + Budrich.
- Sälzer, C., Reiss, K., Schiepe-Tiska, A., Prenzel, M., & Heinze, A. (2013). Zwischen Grundlagenwissen und Anwendungsbezug: Mathematische Kompetenz im internationalen Vergleich. In M. Prenzel, C. Sälzer, E. Klieme & O. Köller (Hrsg.), *PISA 2012: Fortschritte und Herausforderungen in Deutschland* (S. 47–98). Münster: Waxmann.
- Schiepe-Tiska, A., Reiss, K., Obersteiner, A., Heine, J.-H., Seidel, T., & Prenzel, M. (2013). Mathematikunterricht in Deutschland: Befunde aus PISA 2012. In M. Prenzel, C. Sälzer, E. Klieme & O. Köller (Hrsg.), *PISA 2012: Fortschritte und Herausforderungen in Deutschland* (S. 123–154). Münster: Waxmann.
- Schrader, F. W. (1989). *Diagnostische Kompetenzen von Lehrern und ihre Bedeutung für die Gestaltung und Effektivität des Unterrichts*. Frankfurt: Lang.
- Schrader, F. W., & Helmke, A. (2008). Unter der Lupe: Wie in deutschen Klassenzimmern differenziert wird. *Klexer: Magazin für die Grundschule*, 20, 3–5.
- Schrader, F. W., Helmke, A., Hosenfeld, I., Halt, A. C., & Hochweber, J. (2006). Komponenten der Diagnosegenauigkeit von Lehrkräften. Ergebnisse aus Vergleichsarbeiten in der Grundschule. In F. Eder, A. Gastager & F. Hofmann (Hrsg.), *Qualität durch Standards? Beiträge zur 67. AEPF-Tagung in Salzburg* (S. 265–278). Münster: Waxmann.
- Slavin, R. E. (1994). Quality, appropriateness, incentive, and time: A model of instructional effectiveness. *International Journal of Educational Research*, 21(2), 141–157.
- Südkamp, A., Kaiser, J., & Möller, J. (2012). Accuracy of teachers' judgments of students' academic achievement: A meta-analysis. *Journal of Educational Psychology*, 104(3), 743–762.
- Südkamp, A., Möller, J., & Pohlmann, B. (2008). Der Simulierte Klassenraum: Eine experimentelle Untersuchung zur diagnostischen Kompetenz. *Zeitschrift für Pädagogische Psychologie*, 22(3-4), 261–276.
- van Buuren, S., & Oudshoorn, K. (2000). *Multivariate Imputation by Chained Equations: MICE V1.0 User's manual*. Report TNO-PG 00.038. <http://www.stefvanbuuren.nl/publications/MICE%20V1.0%20Manual%20TNO00038%202000.pdf> [01.09.2014].
- van Ophuysen, S. (2010). Professionelle pädagogisch-diagnostische Kompetenz – eine theoretische und empirische Annäherung. In N. Berkemeyer, W. Bos, H. G. Holtappels, N. McElvany

& R. Schulz-Zander (Hrsg.), *Jahrbuch der Schulentwicklung. Daten, Beispiele und Perspektiven* (S. 203–234). Weinheim: Juventa.

Warwas, J., Hertel, S., & Labuhn, A. S. (2011). Bedingungsfaktoren des Einsatzes von adaptiven Unterrichtsformen im Grundschulunterricht. *Zeitschrift für Pädagogik*, 57(6), 854–867.

**Abstract:** The aim of the present study is to examine in how far a differentiating teaching practice in Math and German lessons is linked with the perception of heterogeneity in student performance and with the teachers' diagnostic accuracy. The study is based on performance data from comparative tests (VERA-8) and teacher assessments of  $N = 1803$  students attending 8<sup>th</sup> grade at grammar schools in the federal state of Brandenburg and on student assessments on differentiation in teaching. Hierarchical regression analyses show that, from the students' perspective, Math teachers differentiate the more in their teaching the better they are able to assess the students' performance level. In German lessons, differentiating forms of teaching are less often applied and there are no links with diagnostic accuracy.

**Keywords:** Differentiated Instruction, Diagnostic Competency, Heterogeneity of Abilities, Teaching Quality, Teacher Professionalization

#### **Anschrift der Autor(inn)en**

Dipl.-Psych. Andrea Westphal, Universität Potsdam,  
Humanwissenschaftliche Fakultät, Empirische Unterrichts- und Interventionsforschung,  
Profilbereich Empirische Bildungswissenschaften,  
Karl-Liebknecht-Straße 24/25, 14476 Potsdam, Deutschland  
E-Mail: andrea.westphal@uni-potsdam.de

Dr. Anna Gronostaj, Universität Potsdam,  
Humanwissenschaftliche Fakultät, Empirische Unterrichts- und Interventionsforschung,  
Profilbereich Empirische Bildungswissenschaften,  
Karl-Liebknecht-Straße 24/25, 14476 Potsdam, Deutschland  
E-Mail: anna.gronostaj@uni-potsdam.de

Prof. Dr. Miriam Vock, Humanwissenschaftliche Fakultät,  
Empirische Unterrichts- und Interventionsforschung,  
Profilbereich Empirische Bildungswissenschaften,  
Karl-Liebknecht-Straße 24/25, 14476 Potsdam, Deutschland  
E-Mail: miriam.vock@uni-potsdam.de

Dr. Rico Emmrich, Institut für Schulqualität der Länder Berlin  
und Brandenburg e. V. (ISQ) an der Freien Universität Berlin,  
Otto-von-Simson-Straße 15, 14195 Berlin, Deutschland  
E-Mail: rico.emmrich@isq-bb.de

Peter Harych, Institut für Schulqualität der Länder Berlin  
und Brandenburg e. V. (ISQ) an der Freien Universität Berlin,  
Otto-von-Simson-Straße 15, 14195 Berlin, Deutschland  
E-Mail: peter.harych@isq-bb.de